Multilinguisation d'un éditeur de documents structurés. Application à un dictionnaire trilingue

Huy Khánh PHAN & Christian BOITET
GETA, Institut IMAG
(UJF & CNRS)
BP 53X – 38041 Grenoble Cédex, France
phan@imag.fr, boitet@imag.fr

Résumé

Pour "multilingualiser" (et non simplement "localiser") Grif, un éditeur de documents structurés, nous avons défini un langage de transcription, appelé langage E, analogue aux autres langages (S, P et T) de Grif.

E est utilisé pour compléter la description structurale d'une classe de documents, écrite en S, par une description "linguistique" concernant les systèmes d'écriture utilisés dans les différentes sous-structures des documents de la classe.

Grâce à cette extension multilingue, on a pu construire une première structure de dictionnaire trilingue chinois–français–vietnamien, et l'utiliser sur un dictionnaire réduit.

Mots-clés

Modèle de document, édition de documents structurés, systèmes d'écriture, multilinguisme, transcriptions.

Introduction

En informatique, les problèmes du multilinguisme se posent actuellement avec acuité. Tant les matériels que les logiciels sont hétérogènes et incomplets. Certes, il existe un "documenteur" multilingue, le Star de Xerox [1]. Mais il s'agit d'une solution fermée, sur matériel spécifique.

Quant aux outils dits "multilingues", comme le texteur WinText [16], ou le SGBD 4D, tous deux sur Macintosh, ou encore les diverses extensions du formateur TEX [5, 6, 7, 15], il s'agit en fait de "localisations", qui héritent leurs possibilités et leurs limites du système d'exploitation sous-jacent.

Il est encore impossible, avec ces outils, d'écrire et d'utiliser en même temps deux (ou plus de deux) systèmes d'écriture non "classiques" (chinois, japonais, arabe, par exemple), en sus des systèmes classiques, traitables par simple ajout de polices adéquates.

Pour étudier concrètement et utilement les problèmes de multilinguisation de logiciels, nous avons travaillé sur un "documenteur" (système de production de documents) de haut niveau et très largement paramétrable, Grif, développé par V. Quint et I. Vatton [10, 11, 12].

Nous sommes ainsi arrivés à Grif-m, une version réellement multilingue [9] de Grif, et l'avons utilisé pour construire un prototype de dictionnaire trilingue chinois–français–vietnamien.

Ce type de technique pourrait être appliqué à des "bases lexicales multilingues" plus complexes, comme celles visées par le projet ESPRIT Multilex [8].

I. Multilinguisation de Grif

1.1. Grif et ses langages S, P, T. Introduction de E

Dans Grif, on définit un modèle de document pour décrire une classe de documents à manipuler. Il s'agit de documents structurés. On spécifie séparément la structure logique et la présentation physique.

La structure logique est définie en S et reflète l'organisation arborescente des composants du document, tels que sections, sous-sections, paragraphes, tableaux, formules et sous-formules mathématiques, etc.

La présentation physique est définie en P et décrit l'image concrète de ces composants dans un certain contexte de restitution (écran, imprimante).

On peut définir en Grif tout modèle de document définissable en SGML, avec de plus la possibilité de créer diverses vues (une vue est un filtre donnant une façon de voir certaines parties du document), et d'interagir avec le document, en wysiwyg, à travers toute vue.

Voici la description en S de la structure de la classe "Article". Un article contient, dans l'ordre d'apparition, un titre, des auteurs (au moins 1), des adresses, un résumé, une suite de sections (au moins 2), etc.

```
STRUCTURE Article =
BEGIN
 Titre = TEXT;
 Auteurs  = LIST [1..*] OF
   (Auteur = TEXT);
 Adresses = LIST OF (Adresse=TEXT);
 Résumé   = LIST OF (Paragraphe);
 Suite_Sections = LIST [2..*] OF
   (Section);
END;
Section =
Titre_Section = TEXT;
Suite_Paragr;
Suite_Sous_Sections = LIST [2..*] OF
   Sous_Section
END;
........
```

Figure 1. Un schéma de structure en langage S.

Enfin, le langage T permet d'exprimer la traduction de la représentation "pivot" d'un document Grif en une représentation dans un autre formalisme, tel que SGML, TEX, LATEX, Scribe, ou Troff.

L'une des applications intéressantes de Grif est la construction de classes de documents bilingues (français-anglais, par exemple), voire multilingue.

Dans sa version originale, Grif travaille avec l'alphabet latin de la norme ISO (augmenté de certains caractères accentués usuels) pour le texte, et avec l'alphabet grec pour les formules mathématiques.

Pour multilingualiser Grif, d'une façon cohérente avec sa conception globale, fondée sur la généricité, nous avons défini un langage de "transcription d'entrée", appelé langage E [9]. E est homogène avec les autres langages de Grif. C'est donc un langage descriptif et non impératif.

Grâce à E, on peut décrire en Grif-m de nombreuses caractéristiques de chaque système d'écriture traité, et résoudre ainsi une bonne partie des problèmes posés par le multilinguisme (codage, saisie, restitution et dialogue).

Cependant, les règles de typographie fine (ligatures, forme dépendant du contexte, etc.) ne peuvent pas être décrites en E (il faudrait l'étendre notablement).

1.2. Définition et implémentation du langage E

Pour définir la syntaxe de E, nous avons utilisé la même méta-grammaire que celle utilisée pour définir les langages S, P et T. Au cœur de Grif, on trouve en effet un noyau de générateur de compilateurs.

En E, on définit des paramètres d'édition et des règles de transcription.

Les paramètres concernent les aspects d'entrée-sortie. Par exemple, nous avons des paramètres d'entrée, comme jeux de caractères et méthodes de saisie, et des paramètres de sortie concernant les polices de caractères disponibles pour la restitution et pour le dialogue.

Les règles définissent les représentations utilisées pour la saisie, le codage et la restitution, ainsi que leur correspondance. Une entrée est définie en prenant en compte l'utilisation d'un clavier QWERTY standard. C'est une chaîne de caractères qui dénote la suite de frappes à effectuer.

Dans une règle, une entrée peut donner accès à une ou plusieurs transcriptions, chacune correspondant à un caractère. Un caractère peut avoir également un ou plusieurs codes d'affichage (numéros d'image). Nous avons proposé actuellement trois types de règle de transcription, normal, homophone, et morphologique, correspondant aux relations (1 -> 1 -> 1), (1 -> N -> N) et (1 -> 1 -> N).

Par exemple, A -> 'A' -> 65; est une règle très simple, de type (1 -> 1 -> 1), qui indique que l'entrée A correspond au code interne dénoté par 'A' (#65) et au caractère n° 65 dans les polices disponibles.

La règle ci-dessous permet de saisir l'un des quatre caractères chinois homophones prononcé A au premier ton en entrant a1 au clavier, puis en sélectionnant parmi les possibilités offertes dans un menu.

```
{ commentaire entre accolades }
a1 -> ('A1-7/FU4-2'   -> 1602,{ 阿 }
       'A1-7/FU4-2'   -> 1602,{ 阿 }
       'A1-6/KOU3-3'  -> 6325,{ 啊 }
       'A1-12/JIN1-5' -> 7925,{ 锕 }
       'A1-12/ROU4-4' -> 7571 { 腌 });
```

Figure 2. Une règle de type (1 -> N -> N).

À chaque caractère, nous faisons correspondre son code PS minimal (par rapport au dictionnaire CIHAI [4], de 15000 caractères), présenté plus loin, et son code GB (norme GB 2312-80 [14] pour les caractères chinois simplifiés utilisés en Chine Populaire).

Il y a aussi des règles de type (1 -> 1 -> N), comme la règle dans la figure 3 pour les caractères grecs, où l'entrée s correspond au code interne dénoté par 's' (une transcription "locale" présentée plus loin) et aux caractères σ et ς (variante morphologique de σ quand celui-ci se trouve en fin d'un mot en grec).

```
s -> 's' -> (115 {σ }, 86 { ς });
```

Figure 3. Une règle de type (1 -> 1 -> N).

La syntaxe des règles permet donc de définir un ou plusieurs codages (représentation interne) des caractères d'un système décriture donné. Ainsi, selon l'application visée, on peut utiliser des normes existantes (ASCII "nationaux" ou codages sur deux octets comme JIS ou GB 2312-80), en taille fixe (2, 3, 4 octets) ou variable, ou encore des transcriptions à la TEI (à un caractère correspond une chaîne ASCII "lisible").

Le codage décrit est aussi utilisé pour écrire les intitulés de dialogue. En effet, il faut que l'utilisateur puisse dialoguer avec le système dans la langue de son choix, tout en manipulant un document.

La compilation d'un schéma écrit en E fournit des tables de transcription contenant les caractéristiques décrites dans le schéma. Ces tables, sous forme interne, servent à guider le processus de traitement d'un texte.

Par exemple, pour traiter les caractères avec signes diacritiques vietnamiens, nous produisons une table de vérification qui donne les possibilités de combinaison de deux signes diacritiques, et une table de consultation qui donne les informations de codage et d'affichage des caractères pour le systèmes d'écriture vietnamien.

Notre implémentation du langage E permet aussi la gestion des ressources de restitution. On peut introduire des polices de caractères de styles divers, comme souligné, relief, etc., puis les utiliser dans un schéma écrit en E. Pour l'impression, Grif produit directement un fichier PostScript. On peut aussi utiliser T pour traduire vers un autre format (TEX, LaTeX, SGML...).

Au total, il a suffi de modifier environ 15% du code de Grif pour obtenir Grif-m.

1.3. Réalisation d'une version multilingue de Grif

Une fois E disponible, nous l'avons utilisé pour créer une version multilingue de Grif, permettant de traiter des documents contenant, outre des fragments "classiques", du vietnamien et du chinois.

Cette limitation est uniquement due au délai dont nous disposions. Toutefois, notre réalisation résoud deux problèmes assez fréquents dans les systèmes d'écriture autres que les systèmes fondés sur l'alphabet latin, à savoir la multiplicité des diacritiques (vietnamien, thaï...) et la taille des grands jeux de caractères (chinois, japonais).

Par contre, nous n'avons pas encore traité les problèmes de sens d'écriture, d'analyse de contexte, de ligature..., posés par quelques systèmes d'écriture, comme l'arabe.

Nous avons commencé par écrire en E les schémas de transcription pour le vietnamien et pour le chinois. Pour le vietnamien, il s'agissait de décrire les caractères portant des signes diacritiques (comme pour les caractères accentués latins de la norme ISO), aisin que la méthode de saisie.

Dans la définition suivante (en E), on indique que la saisie se fera par composition à gauche (insertion du ou des diacritiques avant la lettre).

```
PARAMS_IN
Set = Romain;
Language = vietnamien;
Entry = LEFTCOMPO;
 { saisie par composition à gauche }
Letters = (A, D, E, I, O, U, Y, a, d,
           e, i, o, u, y);
Signs = (
 CI ( circonflexe ^
                  (pour A a E e O o)),
 BR ( brève     ˘ (pour A a)      ),
 VQ ( crochet   ' (pour O o U u    )),
 VQ ( grave     `                  ),
 QU ( question  ,                  ),
 TI ( tilde     ~                  ),
 AC ( aigu      '                  ),
 PE ( point     .                  ),
 EN ( barre     -   (pour D d)     ));
Max_Signs = 2;
 { deux signes maximum par lettre }
```

Figure 4. Paramètres d'entrée pour le vietnamien.

Le codage des caractères diacrités vietnamiens (134 au total) utilise une transcription, fondée sur la norme ISO-025 des caractères français, développée et utilisée au GETA [2] pour la TAO d'autre langues utilisant l'alphabet romain.

Dans cette transcription, on utilise la combinaison du signe '!' et d'un nombre pour représenter un diacritique ('!1' pour l'accent aigu ', '!2' pour l'accent grave `, '!3' pour l'accent circonflexe ^, etc.). Par exemple, á est représenté par 'a!1', et ề par 'e!3!2'.

La règle suivante donne le codage du caractère ề avec la correspondance entrée–sortie (entrée par une suite de frappes et numéro d'image dans la police disponible):

```
CI GR e -> 'e!3!2' -> 165;
```

Pour les caractères chinois, nous utilisons le code PS (phonético-structural) de Ch. Boitet et F. Tchéou [3].

Ce code représente un caractère chinois par une suite (minimale) de triplets (pinyin, ton, nombre de traits), le premier concernant le caractère entier, le second sa clé "sémantique" (radical), et les suivants les autres radicaux contenus dans le caractère, par ordre d'écriture.

Il utilise un ensemble réduit de caractères contenu dans l'alphabet de PL/I.

Par exemple, HAO3-6/NU:3-3 représente le caractère chinois 好 ("bon"). Ce caractère se prononce HAO au troisième ton (rentrant ˇ) et comporte 6 traits. Sa clé sémantique (女, "femme") se prononce NU: au troisième ton et a 3 traits.

L'écriture complète d'un schéma de transcription du chinois demande un temps apréciable. Pour accélérer le travail, nous avons utilisé les fichiers de données de saisie pinyin et les polices de caractères chinois simplifiés disponibles au laboratoire GEDIS de l'Université de Lille 1.

II. Application à un dictionnaire trilingue chinois-français-vietnamien

II.1. Description du dictionnaire

Remarquons d'abord qu'il s'agit d'édition structurée, et en aucune façon de gestion de base de données.

D'autre part, la structure que nous définissons pour cette classe de documents (dictionnaires trilingue chinois–français–vietnamien) est volontairement très simple. La rendre plus complexe ne poserait aucun problème.

Un dictionnaire est une liste d'articles rangés dans l'ordre de transcription pinyin [5] des caractères chinois.

Un article contient :

- son numéro.
- l'entrée ou clé (un terme en caractère chinois).
- la prononciation (transcription pinyin suivie d'un chiffre de 1 à 4 désignant le ton, continu ¯, montant ´, rentrant ˇ, descendant `). L'absence de ton représente le ton faible. Par exemple, 大 ("grand") se prononce *dà* et est dénoté *DA4*.
- la catégorie grammaticale (*nom, adj., adv., ver.*).
- le type d'utilisation (normale (*nor.*), familière (*fam.*), argotique (*arg.*), etc.).
- le ou les équivalents français.
- le ou les équivalents vietnamiens.

La vue principale, *Dict_trois_Langues*, présente le contenu complet du dictionnaire.

On définit sa présentation en demandant de mettre en tête le titre en trois langues, suivi de la liste des auteurs et d'un mode d'emploi en français, formé d'une suite de paragraphes.

Les entrées sont rangées en lignes et leurs rubriques en colonnes, dans l'ordre d'apparition.

Les numéros d'entrée sont présentés en caractère gras, les transcriptions phonétiques entre crochets, les abréviations de catégorie et d'utilisation en italiques, et chacun des équivalents français et vietnamiens est affiché sur un paragraphe indépendant.

On a défini trois autres vues. La vue *Vue_Chinoise* est en fait un dictionnaire monolingue montrant uniquement les caractères chinois (graphie, transcription, catégorie et utilisation).

Les vues *Vue_Française* et *Vue_Vietnamienne* sont les dictionnaires bilingues chinois–français et chinois–vietnamien extraits du dictionnaire trilingue.

II.2. Définition en S et P du dictionnaire trilingue

La définition suivante de la structure du dictionnaire, écrite en S, impose qu'il contienne au moins deux entrées (`LISTE [2..*] OF (Entrée = ...)`).

```
STRUCTURE DicTriLing =
BEGIN
  Titre =
  BEGIN
    TitreChinois      = TexteChinois;
    TitreFrançais = TexteFrançais;
    TitreVietnamien   = TexteVietnamien;
  END
  Auteurs = LIST OF
    (Auteur = TexteFrançais);
  ModeEmploi = LIST OF
    (ParagrModeEmploi = TexteFrançais);
  Corps_Dict = LIST [2..*] OF (Entrée =
  BEGIN
    Graphie      = TexteChinois;
    Phonétique   = TexteFrançais;
    Catégorie    = TexteFrançais;
    Utilisation = TexteFrançais;
    Equival_français  = LIST OF
      (Equ_français = TexteFrançais);
    Equival_vietnamien  = LIST OF
      (Equ_français = TexteVietnamien);
  END);
END;
```

Figure 5. Structure en S du dictionnaire trilingue

Aux types Texte_chinois, Texte_français et Texte_vietnamien sont associées les méthodes de saisie adéquates : un élément chinois sera saisi à travers l'interface de saisie du chinois.

Le schéma de présentation en langage P définit la façon de restituer sur l'écran ou sur papier le contenu du dictionnaire en toute vue prévue. On y décrit la présentation physique de chaque élément défini dans le schéma de structure.

Par exemple, pour afficher un caractère chinois sur un rectangle de largeur 1,2 cm et de hauteur celle de la police de caractères chinois disponible, on écrit en P :

```
Graphie :
BEGIN
  Width     : 1.2 cm;
  VertPos   : Top = Enclosing.Top;
  HorizPos  : Left= Enclosing.Left+1 cm;
END;
```

Figure 6. Structure en P d'un élément du dictionnaire trilingue

II.3. Exemple

Voici une partie du dictionnaire trilingue chinois–français–vietnamien édité avec cette version de Grif-m (Sun3 / Unix / X-Window V11R3).

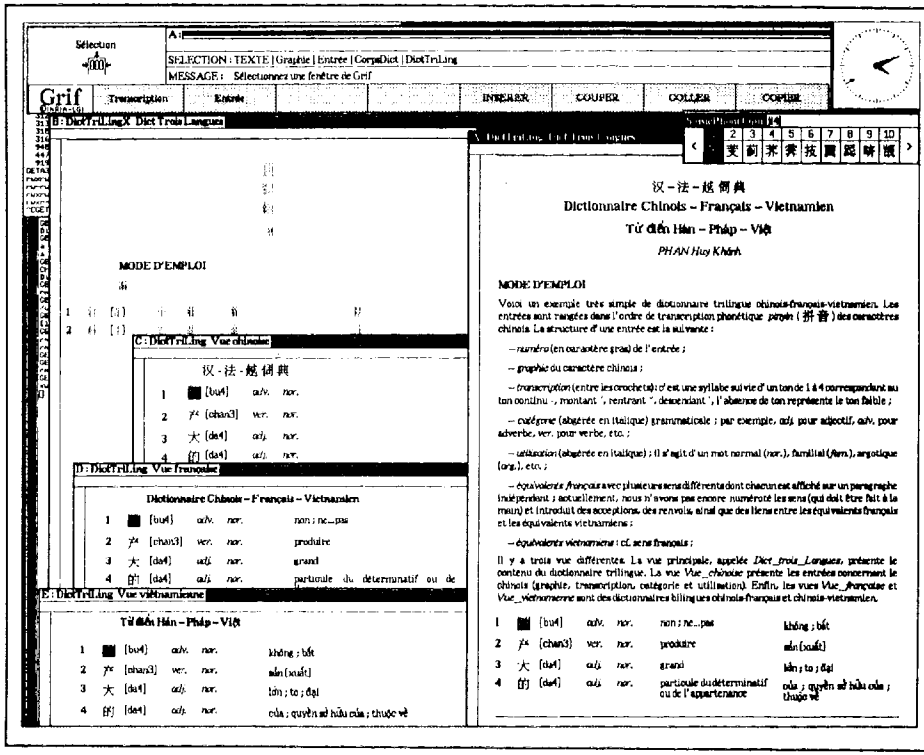

Figure 7. Quatre vues différentes d'un dictionnaire trilingue chinois–français–vietnamien (Dict_trois_Langues, Vue_Chinoise, Vue_Française et Vue_Vietnamienne).

À partir du "squelette", décrit en P par un schéma de présentation, et mis en évidence par une sous-fenêtre contenant des rectangles grisés, l'utilisateur remplit le contenu du dictionnaire dans les langues disponibles. Pour les caractères chinois, il dispose d'une sous-fenêtre de saisie implémentant la méthode de saisie pinyin décrite dans le schéma écrit en E.

III. Perspectives

Ce travail sur la multilinguisation de Grif permet de proposer des méthodes réutilisables dans diverses applications, non seulement en PAO (Publication Assistée par Ordinateur), en EAO (Enseignement Assisté par Ordinateur), mais encore en TALN (Traitement Automatique des Langues Naturelles) : lexicographie, indexation automatique, recherche documentaire, etc.

Par exemple, il devrait être relativement aisé de définir une classe de documents "bitextes", pour obtenir à partir de Grif un éditeur bilingue, avec synchronisation automatique des paragraphes. De même, on pourrait construire une interface conviviale avec des bases lexicales multilingues.

Il reste à compléter les aspects, principalement liés aux problèmes de restitution (césure, typographie fine, sens d'écriture...), pour pouvoir traiter d'autres systèmes d'écriture, comme ceux de l'arabe, de l'hébreu, du thaï, etc.

Ce qui concerne le sens d'écriture est du domaine de la présentation et ne semple pas offrir de difficulté particulière.

Par contre, il faudrait un niveau supplémentaire pour exprimer les règles contextuelles régissant le choix entre différentes formes pour la même lettre (arabe).

Il nous semble également important de travailler dans le cadre de TEI (Text Encoding Initiative) [13], en particulier pour développer des transcriptions utilisables en TEI (où par exemple le signe "!" joue un rôle spécial), ainsi que les transcripteurs associés.

Conclusion

La définition du langage E a permis de compléter la notion de modèle de document de Grif, dans lequel les auteurs de Grif avaient déjà remarquablement bien traité les aspects structuraux et interactifs, au niveau des systèmes d'écriture. L'approche suivie semble nouvelle, et efficace, puisque nous avons réalisé une version prototype réellement multilingue.

Avec cette première version, nous avons construit une structure de dictionnaire trilingue chinois–français–vietnamien assez simple, mais très illustrative, puisqu'on peut manipuler un dictionnaire à travers diverses vues, monolingues, bilingues et trilingues, en traitant naturellement chaque champ dans le système d'écriture approprié.

Remerciements

V. Quint et I. Vatton nous ont permis d'utiliser leur version de recherche de Grif, et n'ont pas ménagé leur temps pour en expliquer les détails. A. Cousquer et son équipe ont mis à notre disposition toutes les resources informatiques nécessaires à la restitution du chinois. Enfin, F. Tchéou nous a grandement aidés pour ce qui concerne la transcription PS des caractères chinois.

Références

[1] J. D. BECKER (1984) *Le traitement de Texte Multilingue* Pour la Science, sept. 1984, 66-76.

[2] C. BOITET, D. BACHUT, R. GERBER (1986) *ARIANE portable : Dossier d'analyse Grands Caractères.*Version 2.0, PN-TAO & GETA, Grenoble, mai 1986, 46 p.

[3] C. BOITET, F. X. TCHÉOU (1990) *On a phonetic and structural encoding of Chinese characters in Chinese texts.* Proc. ROCLING III, Taipeh, Aug. 1990, 71-80.

[4] CIHAI 辭 海 (1983) *Large Dictionary of Chinese Characters and Words* 4th edition, Ci Shu, Shanghai, 1983, 5540 p.

[5] A. COUSQUER (1990) *En chinois dans le TEXte.* Cahiers GUTenberg, n°6, juillet 1990, 15-24.

[6] M. J. FERGUSON (1986) *Multilingual TEX. X for Documentation.*Second European Conference. Strasbourg, June 1986, 19-21.

[7] Y. HARALAMBOUS (1989) *TEX and latin alphabet languages.*TUGboat, 10/3, 1989, 342-345.

[8] G. HEYER, K. WALDHÖR & H. KHATCHADOURIAN (1991)*Motivations, Goals & Milestones of ESPRIT II Project MULTILEX.* Conferences & Exhibition on Language Industry, EC2, Volume 1, Session 10 on R & D, Versailles, Jan. 16-17, 1991.

[9] H. K. PHAN (1991) *Contribution à l'informatique multilingue. Extension d'un éditeur de documents structurés.* Thèse de Doctorat, Université des sciences et techniques de Lille Flandres Artois, mai 1991, 231 p.

[10] V. QUINT, I. VATTON (1986) *Grif: An Interactive System for Structured DocumentManipulation.* Text Processing and Document Manipulation, Proc. of the International Conference, J. C. van Vliet ed., Cambridge University Press, 1986, 200-213.

[11] V. QUINT, I. VATTON, H. BEDOR (1986) *Le système Grif.* T.S.I., 5/4, 1986, 337-341.

[12] V. QUINT (1987) *Une approche de l'édition structurée des documents.* Thèse de Doctorat d'Etat ès-Science Mathématiques, Université Joseph Fourier (Grenoble 1), mai 1987, 283 p.

[13] C. M. SPERBERG-MCQUEEN & L. BURNARD (1990) *ACH–ACL–ALLC Guidelines for the Encoding and Interchange of Machine-Readable Texts.* TEI P1, Draft Version 1.0, Chicago and Oxford, July 1990, 279 p.

[14] The People's Republic of China (1981) *Code of Chinese graphic Character set for Information Interchange. Primary set GB 2312-80.* Fuxing-menwai Sanlihe, Beijing, China, 1981, 175 p.

[15] B. E. VOGEL (1989) *PrintingVietnamese characters by adding diacritical marks via TEX.* TUGboat, 10/2, 1989, 217-223.

[16] WINSOFT (1988) *WinText, le traitement de textes multilingues pour Macintosh.*Version 2.0, WinSoft, 1988, 392 p.

À partir du "squelette", décrit en P par un schéma de présentation, et mis en évidence par une sous-fenêtre contenant des rectangles grisés, l'utilisateur remplit le contenu du dictionnaire dans les langues disponibles. Pour les caractères chinois, il dispose d'une sous-fenêtre de saisie implémentant la méthode de saisie pinyin décrite dans le schéma écrit en E.

III. Perspectives

Ce travail sur la multilinguisation de Grif permet de proposer des méthodes réutilisables dans diverses applications, non seulement en PAO (Publication Assistée par Ordinateur), en EAO (Enseignement Assisté par Ordinateur), mais encore en TALN (Traitement Automatique des Langues Naturelles) : lexicographie, indexation automatique, recherche documentaire, etc.

Par exemple, il devrait être relativement aisé de définir une classe de documents "bitextes", pour obtenir à partir de Grif un éditeur bilingue, avec synchronisation automatique des paragraphes. De même, on pourrait construire une interface conviviale avec des bases lexicales multilingues.

Il reste à compléter les aspects, principalement liés aux problèmes de restitution (césure, typographie fine, sens d'écriture...), pour pouvoir traiter d'autres systèmes d'écriture, comme ceux de l'arabe, de l'hébreu, du thaï, etc.

Ce qui concerne le sens d'écriture est du domaine de la présentation et ne semple pas offrir de difficulté particulière.

Par contre, il faudrait un niveau supplémentaire pour exprimer les règles contextuelles régissant le choix entre différentes formes pour la même lettre (arabe).

Il nous semble également important de travailler dans le cadre de TEI (Text Encoding Initiative) [13], en particulier pour développer des transcriptions utilisables en TEI (où par exemple le signe "!" joue un rôle spécial), ainsi que les transcripteurs associés.

Conclusion

La définition du langage E a permis de compléter la notion de modèle de document de Grif, dans lequel les auteurs de Grif avaient déjà remarquablement bien traité les aspects structuraux et interactifs, au niveau des systèmes d'écriture. L'approche suivie semble nouvelle, et efficace, puisque nous avons réalisé une version prototype réellement multilingue.

Avec cette première version, nous avons construit une structure de dictionnaire trilingue chinois–français–vietnamien assez simple, mais très illustrative, puisqu'on peut manipuler un dictionnaire à travers diverses vues, monolingues, bilingues et trilingues, en traitant naturellement chaque champ dans le système d'écriture approprié.

Remerciements

V. Quint et I. Vatton nous ont permis d'utiliser leur version de recherche de Grif, et n'ont pas ménagé leur temps pour expliquer les détails. A. Cousquer et son équipe ont mis à notre disposition toutes les resources informatiques nécessaires à la restitution du chinois. Enfin, F. Tchéou nous a grandement aidés pour ce qui concerne la transcription PS des caractères chinois.

Références

[1] J. D. BECKER (1984) *Le traitement de Texte Multilingue* Pour la Science, sept. 1984, 66-76.

[2] C. BOITET, D. BACHUT, R. GERBER (1986) *ARIANE portable : Dossier d'analyse Grands Caractères*.Version 2.0, PN-TAO & GETA, Grenoble, mai 1986, 46 p.

[3] C. BOITET, F. X. TCHÉOU (1990) *On a phonetic and structural encoding of Chinese characters in Chinese texts*. Proc. ROCLING III, Taipeh, Aug. 1990, 71-80.

[4] CIHAI 辞海 (1983) *Large Dictionary of Chinese Characters and Words* 4th edition, Ci Shu, Shanghai, 1983, 5540 p.

[5] A. COUSQUER (1990) *En chinois dans le TEX te*. Cahiers GUTenberg, n°6, juillet 1990, 15-24.

[6] M. J. FERGUSON (1986) *Multilingual TEX. X for Documentation*.Second European Conference. Strasbourg, June 1986, 19-21.

[7] Y. HARALAMBOUS (1989) *TEX and latin alphabet languages*.TUGboat, 10/3, 1989, 342-345.

[8] G. HEYER, K. WALDHÖR & H. KHATCHADOURIAN (1991) *Motivations, Goals & Milestones of ESPRIT II Project MULTILEX*. Conferences & Exhibition on Language Industry, EC2, Volume 1, Session 10 on R & D, Versailles, Jan. 16-17, 1991.

[9] H. K. PHAN (1991) *Contribution à l'informatique multilingue. Extension d'un éditeur de documents structurés*. Thèse de Doctorat, Université des sciences et techniques de Lille Flandres Artois, mai 1991, 231 p.

[10] V. QUINT, I. VATTON (1986) *Grif: An Interactive System for Structured DocumentManipulation*. Text Processing and Document Manipulation, Proc. of the International Conference, J. C. van Vliet ed., Cambridge University Press, 1986, 200-213.

[11] V. QUINT, I. VATTON, H. BEDOR (1986) *Le système Grif*. T.S.I., 5/4, 1986, 337-341.

[12] V. QUINT (1987) *Une approche de l'édition structurée des documents*. Thèse de Doctorat d'État ès-Science Mathématiques, Université Joseph Fourier (Grenoble 1), mai 1987, 283 p.

[13] C. M. SPERBERG-MCQUEEN & L. BURNARD (1990) *ACH–ACL–ALLC Guidelines for the Encoding and Interchange of Machine-Readable Texts*. TEI P1, Draft Version 1.0, Chicago and Oxford, July 1990, 279 p.

[14] The People's Republic of China (1981) *Code of Chinese graphic Character set for Information Interchange. Primary set GB 2312-80*. Fuxing-menwai Sanlihe, Beijing, China, 1981, 175 p.

[15] B. E. VOGEL (1989) *Printing Vietnamese characters by adding diacritical marks via TEX*. TUGboat, 10/2, 1989, 217-223.

[16] WINSOFT (1988) *WinText, le traitement de textes multilingues pour Macintosh*.Version 2.0, WinSoft, 1988, 392 p.